PARAMETER-ESTIMATION FOR ARMA MODELS WITH INFINITE VARIANCE INNOVATIONS

Mikosch, Thomas; GADRICH, T; Kluppelberg, Claudia; ADLER, RJ

Published in:
Annals of statistics

DOI:
10.1214/aos/1176324469

The Annals of Statistics
1995, Vol. 23, No. 1, 305–326

PARAMETER ESTIMATION FOR ARMA MODELS WITH INFINITE VARIANCE INNOVATIONS

BY THOMAS MIKOSCH,[1] TAMAR GADRICH, CLAUDIA KLÜPPELBERG AND ROBERT J. ADLER[2]

University of Groningen, Technion, ETH and Technion

We consider a standard ARMA process of the form $\phi(B)X_t = \theta(B)Z_t$, where the innovations Z_t belong to the domain of attraction of a stable law, so that neither the Z_t nor the X_t have a finite variance. Our aim is to estimate the coefficients of ϕ and θ. Since maximum likelihood estimation is not a viable possibility (due to the unknown form of the marginal density of the innovation sequence), we adopt the so-called Whittle estimator, based on the sample periodogram of the X sequence. Despite the fact that the periodogram does not, a priori, seem like a logical object to study in this non-$\mathscr{L}^2$ situation, we show that our estimators are consistent, obtain their asymptotic distributions and show that they converge to the true values faster than in the usual $\mathscr{L}^2$ case.

1. Introduction. Let $\{X_t\}_{t\in\mathbb{Z}}$ be a causal, stationary, autoregressive, moving average [ARMA(p,q)] process satisfying the difference equation

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q} \tag{1.1}$$

in which the innovation sequence $\{Z_t\}_{t\in\mathbb{Z}}$ is a sequence of iid random variables, in the domain of normal attraction of a symmetric stable distribution of unknown index $\alpha \in (0,2)$ and with unknown scaling parameter $\sigma_0 > 0$. The aim of this paper is to find an efficient method for estimating the parameters ϕ_i and θ_j and to derive the asymptotic properties of the estimates. We shall base our estimation on the sample periodogram, as follows:

Let $X_1,\dots,X_n$, $n \ge 1$, be a sample of n observations from an ARMA process of the above kind and define the usual sample periodogram

$$I_{n,X}(\lambda) = \left| n^{-1/\alpha} \sum_{t=1}^{n} X_t e^{-i\lambda t} \right|^2, \qquad -\pi < \lambda \le \pi. \tag{1.2}$$

(The periodogram of $\{Z_t\}_{t\in\mathbb{Z}}$ is defined correspondingly.)

In a setting of Gaussian noise, Whittle (1953) introduced an estimator which now carries his name. The Whittle estimate, $\overline{\beta}_n$, of $\beta =$

Received July 1993; revised February 1994.
[1]Research supported in part by New Zealand FRST Grant 93-VIC-36-039.
[2]Research supported in part by US–Israel Binational Science Foundation Grant 89-298.
AMS 1991 *subject classifications*. Primary 62M10, 62M15; secondary 62E20, 62F10.
Key words and phrases. Stable innovations, ARMA process, periodogram, Whittle estimator, parameter estimation.

$(\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q)^T$ is defined to be the value of β which minimizes the objective function

$$\hat{\sigma}_n^2(\beta) = \frac{1}{n} \sum_j \frac{I_{n,X}(\lambda_j)}{g(\lambda_j, \beta)}, \tag{1.3}$$

where the sum is taken over all Fourier frequencies $\lambda_j = 2\pi j/n \in (-\pi, \pi]$, and the function $g(\lambda, \beta)$ in (1.3) is defined as

$$g(\lambda, \beta) = \frac{|\theta(e^{-i\lambda})|^2}{|\phi(e^{-i\lambda})|^2} = \frac{|1 + \sum_{k=1}^q \theta_k e^{-i\lambda k}|^2}{|1 - \sum_{k=1}^p \phi_k e^{-i\lambda k}|^2}, \qquad -\pi < \lambda \le \pi.$$

Periodograms and spectra are, of course, an intrinsic part of the $\mathscr{L}^2$ structure of stationary stochastic processes. Thus, in the stable setting, where second moments no longer exist, it is not a priori clear what point there is in defining a periodogram at all, let alone using it as a tool for parameter estimation. Nevertheless, there is no intrinsic problem involved in defining it from the data in the usual fashion, and what we shall show in this paper is the result, perhaps unexpected, that the estimator $\bar{\beta}_n$ is a computationally simple estimator of the ARMA parameters and that its asymptotic distribution is remarkably tractable.

One of the main strengths of the Whittle estimator, as is clear from (1.2) and (1.3), is that its value is independent of both α and σ_0. (Note that although α appears in the normalisation defining $\hat{\sigma}_n^2$, this is irrelevant to the minimisation problem.) This is of particular interest, since both of these parameters are difficult to estimate in practice. The stable index α will, however, figure prominently in the asymptotic distribution of the estimators.

The Whittle estimator has been well studied in the case of Gaussian innovations [e.g., Brockwell and Davis (1991), Section 10.8], where it is known to be asymptotically equivalent to the maximum likelihood estimator (MLE). We cannot see how to establish a corresponding result in the present case, nor, in fact, how to find out very much at all about the MLE. The basic problem, of course, is the intractable form of the stable density function. This is not merely a theoretical problem. For example, if one is prepared to assume that the innovation sequence is not only in the domain of attraction of a stable law, but is actually stable, then it seems, at first, to be a straightforward exercise, even without a justifying distributional theory, to attempt numerical (conditional) maximum likelihood estimation. One discovers very rapidly however that all algorithms for calculating stable densities are numerically very delicate (very slowly converging expansion, etc.), so that even from the purely operational side of things, maximum likelihood estimation seems to be a forbidding task. This is not the case with the Whittle estimator, which involves no more numerical difficulties in the stable case than in the Gaussian situation. In fact, from a practical point of view, one can use standard Gaussian time series packages for estimating β, although the confidence intervals they give are no longer valid.

The remainder of the paper is structured as follows: In the following section, we shall present the main results of this paper. These are that $\bar{\beta}_n$ is a consistent (in probability) estimator of β and that it has a characterisable, if not familiar, asymptotic sample distribution. The latter result, of course, yields, *en passant*, the rate of convergence of $\bar{\beta}_n$ to the true parameter values. Perhaps surprisingly, the rate of convergence is considerably better than in the $\mathscr{L}^2$ case. Also in this section we shall introduce two other estimators—one equivalent to the Whittle estimator and one asymptotically equivalent. It is actually for the last of these that we shall derive the appropriate distribution theory.

By way of background, and for comparison purposes, in Section 3 we shall give a brief summary of previously studied estimators for AR(p) processes with innovations in the domain of attraction of a stable law, including the Yule–Walker estimator of Hannan and Kanter (1977), the LAD-estimator of An and Chen (1982) and the M-estimator of Davis, Knight and Liu (1992). It is worth noting, even at this stage, that the results of the current paper stand out in that they are the first time a full ARMA, rather than AR or MA, process has been studied.

Section 4 contains the results of a small simulation study in which we look at the AR(1), MA(1) and ARMA(1, 1) cases. The main feeling that one gains from this study is that our estimator, for the stable case, behaves as well as the usual MLE does for the Gaussian case.

The proofs of the main results appear in the final Section 6, after some preparatory results in Section 5.

2. The Whittle estimator: Main results. As in the preceding section, we consider a causal, invertible ARMA(p, q) process $\{X_t\}_{t \in \mathscr{Z}}$ satisfying

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}$$

for iid $\{Z_t\}_{t \in \mathscr{Z}}$. Set, for complex z with $|z| \le 1$,

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p,$$
$$\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q,$$
$$\psi(z) = \frac{\theta(z)}{\phi(z)} = 1 + \psi_1 z + \psi_2 z^2 + \cdots,$$

so that $\{X_t\}_{t \in \mathscr{Z}}$ also has the infinite moving average representation

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}, \qquad t \in \mathscr{Z}, \tag{2.1}$$

with $\psi_0 = 1$. Denote the autoregressive and moving average parameters together by

$$\beta = (\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q)^T.$$

The natural parameter set for β is

$$C = \{\beta \in \mathscr{R}^{p+q}: \phi_p \ne 0,\ \theta_q \ne 0,\ \phi(z) \text{ and } \theta(z) \text{ have no common zeros, } \phi(z)\theta(z) \ne 0 \text{ for } |z| \le 1\}.$$

We shall consider two sets of assumptions on $\{Z_t\}_{t \in \mathbb{Z}}$. Initially, it suffices to assume that they satisfy the following three assumptions for some $d > 0$ and positive numbers a_n, such that $a_n \uparrow \infty$:

$$\text{(A1)} \qquad E|Z_1|^d < \infty,$$

$$\text{(A2)} \qquad n/a_n^{2\delta} \to 0, \qquad n \to \infty, \quad \text{for } \delta = 1 \wedge d,$$

$$\text{(A3)} \qquad \lim_{x \to 0} \limsup_{n \to \infty} P\left(a_n^{-2} \sum_{t=1}^{n} Z_t^2 \le x \right) = 0.$$

One natural choice for (a_n^2) is given by the unique solution of the equation

$$E\left[1 \wedge \left(\frac{Z_1^2}{x} \right)^2 \right] = \frac{1}{n}, \qquad n \ge 1.$$

Note that (A1) and the exponential decrease of the ψ_j as $j \to \infty$ imply the absolute a.s. convergence of the series (2.1) for every $t \in \mathbb{Z}$. This is a consequence of the three-series theorem. Furthermore, the conditions $EZ_1^2 < \infty$, (A2) and (A3) cannot hold together, since (A2) and the strong law of large numbers imply that

$$a_n^{-2} \sum_{t=1}^{n} Z_t^2 \to 0 \quad \text{a.s.},$$

which contradicts (A3). The final condition (A3) is a relatively weak stochastic compactness condition.

It will turn out that (A1)–(A3) will suffice to establish the consistency of the estimators that we shall propose. However, in order to obtain information on their asymptotic distribution, we require additional assumptions. In particular, we shall assume that Z_1 is in the domain of normal attraction (DNA) of a symmetric, α-stable, random variable $[Z_1 \in \mathrm{DNA}(\alpha)]$ for some (unknown) $\alpha \in (0, 2)$. That is,

$$\text{(2.2)} \qquad n^{-1/\alpha} \sum_{t=1}^{n} Z_t \to_{\mathscr{D}} Y,$$

where Y is symmetric α-stable. Recall that a random variable Y is said to have a stable distribution $[Y =_d S_\alpha(\sigma, \beta, \mu)]$ if there are parameters $0 < \alpha \le 2$, $\sigma \ge 0$, $-1 \le \beta \le 1$ and μ real such that its characteristic function has the form

$$E(e^{itY}) = \begin{cases} \exp\left\{ -\sigma^\alpha |t|^\alpha \left(1 - i\beta (\operatorname{sign} t) \tan \dfrac{\pi\alpha}{2} \right) + i\mu t \right\}, & \text{if } \alpha \ne 1, \\ \exp\left\{ -\sigma |t| \left(1 + \dfrac{2i\beta}{\pi} (\operatorname{sign} t) \ln|t| \right) + i\mu t \right\}, & \text{if } \alpha = 1. \end{cases}$$

If $\beta = \mu = 0$, then Y is symmetric, and we say that Y has a "symmetric α-stable" distribution, denoted by $Y =_{\mathrm{d}} \mathrm{S}\alpha\mathrm{S}$.

Under the above conditions, it is not hard to see that (A1)–(A3) all hold with the somewhat more explicit choice

$$a_n = n^{1/\alpha}, \qquad n \geq 1,$$

for the normalising constants [cf. Feller (1971) or Bingham, Goldie and Teugels (1987)].

We can now return to the estimation problem. Let $g(\lambda, \beta)$ denote the "power transfer function" corresponding to $\beta \in C$; that is,

$$g(\lambda, \beta) = \left| \frac{\theta(e^{-i\lambda})}{\phi(e^{-i\lambda})} \right|^2 = |\psi(e^{-i\lambda})|^2,$$

and denote the self-normalised periodogram by

$$\tilde{I}_{n,X}(\lambda) = \frac{\left|\sum_{t=1}^{n} X_t e^{-i\lambda t}\right|^2}{\sum_{t=1}^{n} X_t^2}, \qquad -\pi < \lambda \leq \pi,$$

where $I_{n,X}$ is the periodogram of the Introduction.

Furthermore, set

$$\sigma_n^2(\beta) = \int_{-\pi}^{\pi} \frac{\tilde{I}_{n,X}(\lambda)}{g(\lambda, \beta)}\, d\lambda, \qquad \bar{\sigma}_n^2(\beta) = \frac{2\pi}{n} \sum_j \frac{\tilde{I}_{n,X}(\lambda_j)}{g(\lambda_j, \beta)},$$

where the sum is taken over all Fourier frequencies $\lambda_j = 2\pi j/n \in (-\pi, \pi]$. Clearly, as $n \to \infty$, the sum and integral here should converge to the same limit.

Suppose $\beta_0 \in C$ is the true, but unknown, parameter vector. Then two natural estimators of β_0 are given by

$$\beta_n = \arg\min_{\beta \in C} \sigma_n^2(\beta), \qquad \bar{\beta}_n = \arg\min_{\beta \in C} \bar{\sigma}_n^2(\beta) = \arg\min_{\beta \in C} \hat{\sigma}_n^2(\beta).$$

Given the assumption that $\sigma_n^2(\beta) \sim \bar{\sigma}_n^2(\beta)$, it seems reasonable to assume, as is in fact the case, that $\beta_n \sim \bar{\beta}_n$, and that therefore the two estimators are asymptotically equivalent. It is clear that, in practice, $\bar{\beta}_n$ is the only applicable estimator, since the integral defining $\sigma_n^2(\beta)$ will always have to be evaluated by an approximating sum. Nevertheless, throughout this paper we shall give proofs of convergence for the estimator based on $\sigma_n^2(\beta)$, since here the notation is considerably lighter. A parallel proof for $\bar{\beta}_n$ (based on the function $\hat{\sigma}_n^2$) can be found in Gadrich (1993).

The choice of these estimators is motivated by two facts. The first is that the function

$$\int_{-\pi}^{\pi} \frac{g(\lambda, \beta_0)}{g(\lambda, \beta)}\, d\lambda$$

has its absolute minimum at $\beta = \beta_0$ in C [cf. Brockwell and Davis (1991), Proposition 10.8.1]. Moreover, by previous results of Klüppelberg and Mikosch

(1994), $\tilde{I}_{n,X}(\lambda)$ can be applied to estimate $g(\lambda, \beta_0)/\Psi^2(\beta_0)$, where $\Psi^2(\beta_0)$ is the quantity

$$\Psi^2(\beta_0) = \sum_{j=0}^{\infty} \psi_j^2 \tag{2.3}$$

corresponding to β_0.

Of equal importance, however, is the fact that in the Gaussian case the estimator β_n is closely related to least squares and maximum likelihood estimators and is therefore a standard estimator for ARMA processes with finite variance. The idea goes back to Whittle (1953), with a rigorous derivation of the corresponding sample distribution due to Hannan (1973). [See also Fox and Taqqu (1986) and Dahlhaus (1989).] It is well known that in the classical case β_n is consistent and asymptotically normal [cf. Brockwell and Davis (1991)]. The fact that the same estimator works in both the Gaussian and infinite variance cases is a strong argument in favour of its use.

Our first, consistency, result is the following theorem.

THEOREM 2.1. *Suppose* $\{X_t\}_{t\in\mathbb{Z}}$ *is a causal, invertible,* ARMA(p,q) *process and conditions* (A1)–(A3) *hold. Then*

$$\beta_n \to_{\mathscr{P}} \beta_0 \quad \text{and} \quad \sigma_n^2(\beta_n) \to_{\mathscr{P}} 2\pi\Psi^{-2}(\beta_0), \qquad n\to\infty.$$

Furthermore, the same limit relationships hold also for $\bar{\beta}_n$ *and* $\bar{\sigma}_n^2$.

As an interesting aside, we note that in the Gaussian case a corresponding result holds for the quantity corresponding to $\hat{\sigma}_n^2(\bar{\beta}_n)$, which converges to the innovation variance σ_0^2. That is, in the Gaussian case, the self-normalisation in the definition of $\bar{\sigma}_n^2$ is not required. In the current, non-$\mathscr{L}^2$, situation, $\hat{\sigma}^2(\bar{\beta}_n)$ actually converges, in distribution, to a random variable. It is this fact, which has a considerable complicating influence on the proof of all the results related to $\hat{\sigma}^2(\bar{\beta}_n)$, that makes us prefer the self-normalised, but numerically equivalent, estimator based on $\bar{\sigma}_n^2$.

For ARMA(p,q) processes with finite variance, β_n is asymptotically normal with rate of convergence of order $\sqrt{n}$. An analogous result in our case gives a rather less familiar asymptotic distribution, but a considerably faster rate of convergence of order $(n/\ln n)^{1/\alpha}$ (recall that $\alpha < 2$). To state this result, let C_α be the constant defined by

$$C_\alpha = \begin{cases} \dfrac{1-\alpha}{\Gamma(2-\alpha)\cos(\pi\alpha/2)}, & \text{if } \alpha \neq 1,\\[2ex] \dfrac{2}{\pi}, & \text{if } \alpha = 1.\end{cases}$$

THEOREM 2.2. *Suppose* $\{X_t\}_{t\in\mathbb{Z}}$ *is an* ARMA(p,q) *process and* $\{Z_t\}_{t\in\mathbb{Z}}$ *are iid symmetric such that* (2.2) *holds for some* $\alpha < 2$. *Then*

$$\left(\frac{n}{\ln n}\right)^{1/\alpha}(\beta_n - \beta_0) \to_{\mathscr{D}} 4\pi W^{-1}(\beta_0)\frac{1}{Y_0}\sum_{k=1}^{\infty} Y_k b_k, \tag{2.4}$$

where $Y_0, Y_1, Y_2, \ldots$ *are independent random variables,* $Y_0 =_d S_{\alpha/2}(C_{\alpha/2}^{-2/\alpha}, 1, 0)$ *is positive* $\alpha/2$*-stable,* $(Y_t)_{t \in \mathscr{N}}$ *are iid* $S\alpha S$ *with scale parameter* $\sigma = C_\alpha^{1/\alpha}$, $W^{-1}(\beta_0)$ *is the inverse of the matrix*

$$W(\beta_0) = \int_{-\pi}^{\pi} \left[\frac{\partial \ln g(\lambda, \beta_0)}{\partial \beta}\right]\left[\frac{\partial \ln g(\lambda, \beta_0)}{\partial \beta}\right]^T d\lambda,$$

and, for $k \in \mathscr{N}$, b_k *is the vector*

$$b_k = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-ik\lambda} g(\lambda, \beta_0) \frac{\partial g^{-1}(\lambda, \beta_0)}{\partial \beta}\, d\lambda, \tag{2.5}$$

where g^{-1} *denotes the reciprocal of* g*. Furthermore,* (2.4) *holds also with* β_n *replaced by* $\bar{\beta}_n$.

Note that the limit vector in (2.4) is the ratio of an α-stable $(p+q)$-dimensional vector over a positive $\alpha/2$-stable random variable.

3. On what has gone before. There is a small, but interesting and rapidly growing, literature on parametric estimation for ARMA processes with infinite variance innovations. As with our approach, the difficulties in developing a maximum likelihood estimator have led to a number of essentially ad hoc procedures, each of which generalizes some aspect of the Gaussian case. Nevertheless, a relatively consistent picture, at least as far as rates of convergence are concerned, has developed. Not surprisingly, the first estimator studied was a Yule–Walker (YW) type estimate for the parameters of an AR(p) process.

The YW estimates $\hat{\phi}_{\mathrm{YW}}$ of the true values ϕ_0 of an AR(p) process are defined as the solution of

$$\tilde{\Gamma}\phi_{\mathrm{YW}} = \tilde{\gamma},$$

where $\tilde{\Gamma} = [\tilde{\gamma}_n(i-j)]_{i,j=1}^p$, $\tilde{\gamma} = (\tilde{\gamma}_n(1), \ldots, \tilde{\gamma}_n(p))^T$ and $\tilde{\gamma}_n(h)$ is the sample autocorrelation function, $\tilde{\gamma}_n(h) = C(h)/C(0)$, $h \geq 0$, where $C(h) = n^{-1}\sum_{t=1}^{n-h} X_t X_{t+h}$. In the autoregressive case it is not difficult to see that the YW estimate coincides with the Whittle estimate based on σ_n^2.

Hannan and Kanter (1977) showed that if $0 < \alpha < 2$ and $\delta > \alpha$, then

$$n^{1/\delta}\left(\hat{\phi}_{\mathrm{YW}} - \phi_0\right) \to 0 \quad \text{a.s. as } n \to \infty.$$

More recently, Davis and Resnick (1986) showed that there exists a slowly varying function $L_0(n)$ such that

$$n^{1/\alpha}L_0(n)\left(\hat{\phi}_{\mathrm{YW}} - \phi_0\right) \to_{\mathscr{D}} \mathbf{Y} \quad \text{as } n \to \infty,$$

where the structure of $\mathbf{Y}$ is closely related to the right-hand side of (2.4).

A somewhat different approach to parameter estimation, still in the purely autoregressive case, is based on a least absolute deviation (LAD) estimator, which we denote by $\hat{\phi}_{\mathrm{LAD}}$. The LAD estimate of ϕ_0 is defined as the minimizer of

$$\sum_{t=1}^{n} |X_t - \varphi_1 X_{t-1} - \cdots - \varphi_p X_{t-p}|$$

with respect to $\varphi = (\varphi_1, \dots, \varphi_p)^T$.

An and Chen (1982) showed that if Z_1 has a unique median at zero and Z_1 is in the domain of attraction of a stable distribution with index $\alpha \in (1,2)$, or Z_1 has a Cauchy distribution centered at zero, then, for $\delta > \alpha$,

$$n^{1/\delta}\big(\hat{\phi}_{\mathrm{LAD}} - \phi_0\big) \to_{\mathscr{P}} 0, \quad \text{as } n \to \infty.$$

More recently, Davis, Knight and Liu (1992) defined the M-estimate $\hat{\phi}_M$ of an AR(p) process as the minimizer of the objective function

$$U_n(\varphi) = \sum_{t=p+1}^{n} \rho\big(X_t - \varphi_1 X_{t-1} - \cdots - \varphi_p X_{t-p}\big)$$

with respect to φ, where $\rho(\cdot)$ is some loss function. They also established the weak convergence of $\hat{\phi}_M$, for the case when ρ is convex with a Lipschitz continuous derivative. Specifically, they showed that

$$n^{1/\alpha} L_1(n)\big(\hat{\phi}_M - \phi_0\big) \to_{\mathscr{D}} \xi \quad \text{as } n \to \infty,$$

where ξ is the position of the minimum of a certain random field and $L_1(x)$ is a certain slowly varying function. Their analysis seems extendable, without too much effort, to a full ARMA model.

Thus, as is the case for the Whittle estimator, the rate of convergence of the estimator is better than that in the Gaussian case, while the asymptotic distribution is considerably less familiar.

In closing we note that "more rapid than Gaussian" rates of convergence for estimators in heavy-tailed problems seem to be the norm rather than the exception. For example, Feigin and Resnick (1992, 1994) study parameter estimation for autoregressive processes with *positive*, heavy-tailed innovations, and obtain rates of convergence for their estimator of the same order as ours, but without the slowly varying term. Their estimators, however, are different from ours both in spirit and detail, and involve the numerical solution of a nontrivial linear programming problem. Finally, Hsing [(1993), Theorem 3.1] suggests an estimator based on extreme value considerations, which work for the pure AR case. Once again, he obtains an asymptotic distribution reminiscent of (2.4), with a similar rate of convergence.

4. An application to simulated data. To get some idea of how the Whittle estimator behaves in the heavy-tailed situation, we ran a small simulation study using the estimator $\overline{\beta}_n$ based on the summed periodogram $\hat{\sigma}_n^2$.

It should be emphasized that the estimation requires knowledge of neither the stability parameter α nor the scale parameter σ_0 of the data.

TABLE 1
Estimating the parameters of stable and normal ARMA processes via Whittle and MLE estimates

Model No.	True Values	Whittle Estimate		Maximum Likelihood	
		Mean	St. Dev.	Mean	St. Dev.
1	$\phi = 0.4$	0.384	0.093	0.394	0.102
2	$\theta = 0.8$	0.782	0.097	0.831	0.099
3	$\phi = 0.4$	0.397	0.100	0.385	0.106
	$\theta = 0.8$	0.736	0.124	0.815	0.082

Table 1 summarizes some of our results. We generated 100 observations from each of the models:

1. $X_t - 0.4\ X_{t-1} = Z_t$,
2. $X_t = Z_t + 0.8\ Z_{t-1}$,
3. $X_t - 0.4\ X_{t-1} = Z_t + 0.8\ Z_{t-1}$,

where the innovations sequence $\{Z_t\}_{t \in \mathscr{Z}}$ was either iid α-stable with $\alpha = 1.5$ and scale parameter equal to 2.0, or, for comparison purposes, $N(0, 2)$. [In the stable case we relied on the algorithm given by Chambers, Mallows and Stuck (1976) for generation of the innovation process.] We ran 1000 such simulations for each model. In the stable example we estimated the ARMA parameters via the estimator $\bar{\beta}_n$ and in the Gaussian case, via the usual MLE estimator. The results are given in Table 1.

We shall not attempt to interpret these results for the reader, but merely point out that the accuracy of the Whittle estimator in the stable case seems indistinguishable from that of the MLE in the Gaussian case.

Finally, a comment about estimating p, q, α and the scale parameter of the stable innovations. We have assumed throughout, including in the simulation above, that p and q are known. When this is not the case, Bhansali (1984, 1988) and Knight (1989) have proposed techniques for estimating p and q that seem to work well in practice. Estimation of α can be done either from the raw data or from the residuals calculated after parameter estimation. Limited experience with simulations indicates that it is best done on the (supposedly iid) residual. There are a number of techniques available, including those due to Hahn and Weiner (1991), Hsing (1991), McCulloch (1986) and Koutrouvelis (1980). We have found McCulloch's technique to be the most efficient and accurate.

5. Auxiliary results. Since the results of this section hold in somewhat more generality than those of the remainder of the paper, we consider now the more general linear process

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}, \qquad t \in \mathscr{Z}.$$

We suppose the conditions (A1)–(A3) and, additionally, that

$$\sum_{j=-\infty}^{\infty} |\psi_j|^{\delta} |j| < \infty, \qquad \delta = 1 \wedge d.$$

For notational ease we use the same symbol for the periodogram in this more general situation:

$$I_{n,X}(\lambda) = a_n^{-2} \left| \sum_{t=1}^{n} X_t e^{-i\lambda t} \right|^2, \qquad -\pi < \lambda \le \pi,$$

and define $I_{n,Z}(\lambda)$ correspondingly. Note that we chose $a_n = n^{1/\alpha}$ for $Z_1 \in \mathrm{DNA}(\alpha)$ so that this notation is consistent with (1.2). For $h \in \mathscr{Z}$ define

$$\tilde{\gamma}_{n,X}(h) = \gamma_{n,X}(h)/\gamma_{n,X}^2, \qquad \tilde{\gamma}(h) = \gamma(h)/\Psi^2(\beta_0),$$

where $\Psi^2(\beta_0) = \sum_{j=-\infty}^{\infty} \psi_j^2$ [for notational ease we use the same symbol for (2.3) in this more general situation] and

$$\gamma_{n,X}(h) = a_n^{-2} \sum_{t=1}^{n-|h|} X_t X_{t+|h|},$$

$$\gamma(h) = \sum_{j=-\infty}^{\infty} \psi_j \psi_{j+|h|},$$

$$\gamma_{n,X}^2 = a_n^{-2} \sum_{t=1}^{n} X_t^2.$$

The quantities $\gamma_{n,Z}^2$, $\gamma_{n,Z}(h)$ and $\tilde{\gamma}_{n,Z}(h)$ are defined correspondingly. Obviously, if $EZ_1^2 < \infty$, $\tilde{\gamma}_{n,X}(h)$ is a consistent estimator of the autocorrelation function $\tilde{\gamma}(h)$ of $\{X_t\}_{t \in \mathscr{Z}}$. For $Z_1 \in \mathrm{DNA}(\alpha)$, $\alpha \in (0,2)$, Davis and Resnick (1986) showed that $\tilde{\gamma}_{n,X}(h)$ is consistent in probability with limit $\tilde{\gamma}(h)$ and rate of convergence $(n/\ln n)^{1/\alpha}$.

We shall frequently make use of the following decomposition of the periodogram. Its proof is analogous to Proposition 2.1 of Klüppelberg and Mikosch (1993).

PROPOSITION 5.1. *Under the above conditions,*

$$I_{n,X}(\lambda) = |\psi(e^{-i\lambda})|^2 I_{n,Z}(\lambda) + R_n(\lambda), \qquad -\pi < \lambda \le \pi,$$

where

$$R_n(\lambda) = \psi(e^{-i\lambda}) J_n(\lambda) Y_n(-\lambda) + \psi(e^{i\lambda}) J_n(-\lambda) Y_n(\lambda) + |Y_n(\lambda)|^2,$$

$$J_n(\lambda) = a_n^{-1} \sum_{t=1}^{n} Z_t e^{-i\lambda t},$$

$$Y_n(\lambda) = a_n^{-1} \sum_{j=-\infty}^{\infty} \psi_j e^{-i\lambda j} U_{nj}(\lambda),$$

$$U_{nj}(\lambda) = \sum_{t=1-j}^{n-j} Z_t e^{-i\lambda t} - \sum_{t=1}^{n} Z_t e^{-i\lambda t}.$$

The proof of the following lemma is analogous to Davis and Resnick [(1986), page 549]. See also Lemma 5.1 of Klüppelberg and Mikosch (1994).

LEMMA 5.2. *Under the above conditions,*

$$\gamma_{n,X}^2 = \Psi^2(\beta_0)\gamma_{n,Z}^2(1 + o_P(1)), \qquad n \to \infty.$$

PROPOSITION 5.3. *Under the above conditions,*

$$\tilde{\gamma}_{n,X}(h) \to_{\mathscr{P}} \tilde{\gamma}(h), \qquad h \in \mathscr{N}, \quad n \to \infty.$$

PROOF. We mimic the proof of Davis and Resnick [(1986), pages 548–550]. For any $h \in \mathscr{N}$ we have

$$\begin{aligned}\sum_{t=1}^{n} X_t X_{t+h} - \tilde{\gamma}(h) \sum_{t=1}^{n} X_t^2 &= \sum_{t=1}^{n} \sum_{i \neq j} \psi_i\big(\psi_{j+h} - \tilde{\gamma}(h)\psi_j\big) Z_{t-i} Z_{t-j} \\ &\quad + \sum_{t=1}^{n} \sum_{i} \psi_i\big(\psi_{i+h} - \tilde{\gamma}(h)\psi_i\big)\big(Z_{t-i}^2 - Z_t^2\big) \\ &:= V_1 + V_2,\end{aligned} \tag{5.1}$$

where we used the fact that $\sum_i \psi_i \ (\psi_{i+h} - \tilde{\gamma}(h)\psi_i) = 0$. We obtain for some $c_i > 0$, $i = 1,2,3,4$,

$$\begin{aligned}E|a_n^{-2}V_1|^\delta &\le c_1 n a_n^{-2\delta} \sum_{i \neq j} \big|\psi_i\big(\psi_{j+h} - \tilde{\gamma}(h)\psi_j\big)\big|^\delta \\ &\le c_2 n a_n^{-2\delta} \to 0,\end{aligned}$$

and, by (A2),

$$\begin{aligned}E|a_n^{-2}V_2|^{\delta/2} &\le c_3 a_n^{-\delta} \sum_{i} \big|\psi_i\big(\psi_{i+h} - \tilde{\gamma}(h)\psi_i\big)\big|^{\delta/2} |i| \\ &\le c_4 a_n^{-\delta} \to 0.\end{aligned}$$

By Markov's inequality this implies

$$a_n^{-2}(V_1 + V_2) \to_{\mathscr{P}} 0. \tag{5.2}$$

Furthermore, by Lemma 5.2,

$$\gamma_{n,X}^2 = \Psi^2(\beta_0)\gamma_{n,Z}^2(1 + o_P(1)).$$

This, (5.1), (5.2) and (A3) imply that

$$\begin{aligned}\tilde{\gamma}_{n,X}(h) - \tilde{\gamma}(h) &= \frac{\sum_{t=1}^n X_t X_{t+h} - \tilde{\gamma}(h)\sum_{t=1}^n X_t^2}{\sum_{t=1}^n X_t^2} - \frac{\sum_{t=n-h+1}^n X_t X_{t+h}}{\sum_{t=1}^n X_t^2} \\ &= \frac{V_1 + V_2}{\sum_{t=1}^n X_t^2} - \frac{\sum_{t=n-h+1}^n X_t X_{t+h}}{\sum_{t=1}^n X_t^2} = o_P(1).\end{aligned}$$

In the latter relation we also used (A3) together with the fact that $a_n^{-2}\sum_{t=n-h+1}^n X_t X_{t+h} \to_{\mathscr{P}} 0$ for every h. □

6. Proofs of the results in Section 2. The proofs in this section are modelled on those in the finite variance case, due initially to Hannan (1973) [cf. the treatment in Brockwell and Davis (1991), Section 10.8, which we follow closely]. The technical differences in the infinite variance case are, however, substantial. We start with some auxiliary results.

LEMMA 6.1. *Suppose $\{X_t\}_{t\in\mathscr{Z}}$ satisfies the condtions of Theorem 2.1. Then for every $\beta \in C$,*

$$\sigma_n^2(\beta) \to_{\mathscr{P}} \Psi^{-2}(\beta_0)\int_{-\pi}^{\pi}\frac{g(\lambda,\beta_0)}{g(\lambda,\beta)}\,d\lambda, \tag{6.1}$$

and for every $\delta > 0$,

$$\sup_{\beta\in\overline{C}}\left|\sigma_{n,\delta}^2(\beta) - \Psi^{-2}(\beta_0)\int_{-\pi}^{\pi}\frac{g(\lambda,\beta_0)}{g_\delta(\lambda,\beta)}\,d\lambda\right| \to_{\mathscr{P}} 0, \tag{6.2}$$

where $\overline{C}$ denotes the closure of C,

$$g_\delta(\lambda,\beta) = \frac{|\theta(e^{-i\lambda})|^2 + \delta}{|\phi(e^{-i\lambda})|^2}$$

and

$$\sigma_{n,\delta}^2(\beta) = \int_{-\pi}^{\pi}\frac{\tilde{I}_{n,X}(\lambda)}{g_\delta(\lambda,\beta)}\,d\lambda.$$

PROOF. We shall only show that (6.2) is satisfied. The proof of (6.1) is analogous. Set

$$q_m(\lambda,\beta) = \frac{1}{m}\sum_{j=0}^{m-1}\sum_{|k|\le j} r_k e^{-i\lambda k} = \sum_{|k|<m}\left(1-\frac{|k|}{m}\right) r_k e^{-i\lambda k},$$

where

$$r_k = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{i\lambda k} g_\delta^{-1}(\lambda,\beta)\,d\lambda.$$

Fix $\varepsilon > 0$. Then there exists some $m \in \mathscr{N}$ such that

$$|q_m(\lambda,\beta) - g_\delta^{-1}(\lambda,\beta)| < \varepsilon/(4\pi)$$

for all $(\lambda, \beta) \in [-\pi, \pi] \times \overline{C}$. Hence

$$\left|\sigma_{n,\delta}^2(\beta) - \int_{-\pi}^{\pi} \tilde{I}_{n,X}(\lambda) q_m(\lambda, \beta)\, d\lambda\right| \le \frac{\varepsilon}{4\pi} \int_{-\pi}^{\pi} \tilde{I}_{n,X}(\lambda)\, d\lambda = \varepsilon/2, \quad \forall \beta \in \overline{C}.$$

Hence, for fixed ε,

$$\begin{aligned}
&P\left(\sup_{\beta \in \overline{C}} \left|\sigma_{n,\delta}^2(\beta) - \Psi^{-2}(\beta_0) \int_{-\pi}^{\pi} \frac{g(\lambda, \beta_0)}{g_\delta(\lambda, \beta)}\, d\lambda\right| \ge \varepsilon\right) \\
&\quad \le P\left(\sup_{\beta \in \overline{C}} \left|\int_{-\pi}^{\pi} \tilde{I}_{n,X}(\lambda) q_m(\lambda, \beta)\, d\lambda - \Psi^{-2}(\beta_0) \int_{-\pi}^{\pi} \frac{g(\lambda, \beta_0)}{g_\delta(\lambda, \beta)}\, d\lambda\right| \ge \frac{\varepsilon}{2}\right) \\
&\quad = P\left(\sup_{\beta \in \overline{C}} \left|2\pi \sum_{|h| \le m} \tilde{\gamma}_{n,X}(h)\left(1 - \frac{|h|}{m}\right) r_h - \Psi^{-2}(\beta_0) \int_{-\pi}^{\pi} \frac{g(\lambda, \beta_0)}{g_\delta(\lambda, \beta)}\, d\lambda\right| \ge \frac{\varepsilon}{2}\right) \\
&\quad \le P\left(\sup_{\beta \in \overline{C}} \left|2\pi \sum_{|h| \le m} (\tilde{\gamma}_{n,X}(h) - \tilde{\gamma}(h))\left(1 - \frac{|h|}{m}\right) r_h\right| \ge \frac{\varepsilon}{4}\right) \\
&\qquad + I_{[\varepsilon/4, \infty)}\left(\sup_{\beta \in \overline{C}} \left|2\pi \sum_{|h| \le m} \tilde{\gamma}(h)\left(1 - \frac{|h|}{m}\right) r_h - \Psi^{-2}(\beta_0) \int_{-\pi}^{\pi} \frac{g(\lambda, \beta_0)}{g_\delta(\lambda, \beta)}\, d\lambda\right|\right),
\end{aligned}$$

where $I_A(x)$ denotes the indicator function of the set A. The first summand on the right-hand side converges to zero in view of Proposition 5.3 and as a result of the fact that r_h are uniformly bounded for $\beta \in \overline{C}$ and m fixed. The second summand is zero provided m is chosen sufficiently large. □

PROOF OF THEOREM 2.1. We suppose that β_n does not converge in probability to β_0 and obtain a contradiction. From the definition of β_n, we have that, for every t,

$$P(\sigma_n^2(\beta_n) \le t) \ge P(\sigma_n^2(\beta_0) \le t) \to I_{[0,t]}(2\pi \Psi^{-2}(\beta_0)), \tag{6.3}$$

where the convergence is a consequence of Lemma 6.1. By the Helly–Bray theorem and the compactness of $\overline{C}$, there exists a nonrandom subsequence n_k such that β_{n_k} converges in distribution to a random variable β which is different from β_0 on a set of positive probability. The functional $F(f, x) = f(x)$ mapping $\mathscr{C}(\overline{C}) \times \overline{C}$ to $\mathscr{R}$ is continuous, where $\mathscr{C}(\overline{C})$ is the space of continuous functions on $\overline{C}$ equipped with the sup norm. According to Lemma 6.1, $\sigma_{n,\delta}^2(\cdot)$ converges in probability to

$$\Psi^{-2}(\beta_0) \int_{-\pi}^{\pi} \frac{g(\lambda, \beta_0)}{g_\delta(\lambda, \cdot)}\, d\lambda.$$

Hence $\sigma_{n,\delta}^2$ is tight. Since $\beta_{n_k} \to_{\mathscr{D}} \beta$ the sequence β_{n_k} is tight as well. Thus $(\sigma_{n_k,\delta}^2, \beta_{n_k})$ is tight in $\mathscr{C}(\overline{C}) \times \overline{C}$ and so there exists a further subsequence (we continue to use n_k for ease of notation) such that $(\sigma^2_{n_k,\delta}, \beta_{n_k})$ converges in distribution. By the continuous mapping theorem, we conclude that

$$F\big(\sigma^2_{n_k,\delta}, \beta_{n_k}\big) = \sigma^2_{n_k,\delta}\big(\beta_{n_k}\big) \to_{\mathscr{D}} \Psi^{-2}(\beta_0)\int_{-\pi}^{\pi} \frac{g(\lambda,\beta_0)}{g_\delta(\lambda,\beta)}\,d\lambda := \Psi^{-2}(\beta_0)T_\delta(\beta).$$

For a continuity point $t = (2\pi + \varepsilon)\Psi^{-2}(\beta_0)$, $\varepsilon > 0$, of $T_\delta(\beta)$ we have that

$$\limsup_{k\to\infty} P\big(\sigma^2_{n_k}(\beta_{n_k}) \le t\big) \le P(\beta = \beta_0) + P(T_\delta(\beta) \le 2\pi + \varepsilon, \beta \ne \beta_0).$$

Now, letting δ tend to zero, we have

$$\limsup_{k\to\infty} P\big(\sigma^2_{n_k}(\beta_{n_k}) \le t\big) \le P(\beta = \beta_0) + P(T_0(\beta) \le 2\pi + \varepsilon, \beta \ne \beta_0),$$

where

$$T_0(\beta) = \int_{-\pi}^{\pi} \frac{g(\lambda,\beta_0)}{g(\lambda,\beta)}\,d\lambda$$

may assume the value ∞ at the boundary of $\overline{C}$. Note that

$$\{\beta: T_0(\beta) > 2\pi\} \cap \overline{C} = \{\beta: \beta \ne \beta_0\} \cap \overline{C}$$

[cf. Proposition 10.8.1 in Brockwell and Davis (1991)]. By (6.3), we conclude that for sufficiently small ε and t of the above form,

$$\begin{aligned} 1 &= \limsup_{k\to\infty} P\big(\sigma^2_{n_k}(\beta_{n_k}) \le t\big) \\ &\le P(\beta = \beta_0) + P(2\pi < T_0(\beta) \le 2\pi + \varepsilon, \beta \ne \beta_0) \\ &= P(\beta = \beta_0) + P(2\pi < T_0(\beta) \le 2\pi + \varepsilon) \end{aligned}$$

and the right-hand side can be made arbitrarily close to $P(\beta = \beta_0) < 1$ which yields a contradiction and proves the theorem. □

LEMMA 6.2. *Suppose the assumptions of Theorem* 2.2 *hold. Furthermore, let* η *be a continuous function on* $[-\pi, \pi]$ *such that*

$$\left(\frac{n}{\ln n}\right)^{1/\alpha} \int_{-\pi}^{\pi} \eta(\lambda) g(\lambda, \beta_0) \tilde{I}_{n,Z}(\lambda)\,d\lambda = O_p(1), \qquad n \to \infty.$$

Then

$$\left(\frac{n}{\ln n}\right)^{1/\alpha} \int_{-\pi}^{\pi} \big(\tilde{I}_{n,X}(\lambda) - \Psi^{-2}(\beta_0) g(\lambda,\beta_0)\tilde{I}_{n,Z}(\lambda)\big)\eta(\lambda)\,d\lambda \to_{\mathscr{P}} 0.$$

PROOF. Set $x_n = (n/\ln n)^{1/\alpha}$ and recall that $a_n = n^{1/\alpha}$. By Proposition 5.1 and Lemma 5.2 we get

$$\begin{aligned} \int_{-\pi}^{\pi} \tilde{I}_{n,X}(\lambda)\eta(\lambda)\,d\lambda &= \Psi^{-2}(\beta_0)(1 + o_P(1))\int_{-\pi}^{\pi} \gamma_{n,Z}^{-2} I_{n,X}(\lambda)\eta(\lambda)\,d\lambda \\ &= \Psi^{-2}(\beta_0)(1 + o_P(1))\int_{-\pi}^{\pi} \tilde{I}_{n,Z}(\lambda) g(\lambda,\beta_0)\eta(\lambda)\,d\lambda \\ &\quad + \Psi^{-2}(\beta_0)(1 + o_P(1))\gamma_{n,Z}^{-2}\int_{-\pi}^{\pi} R_n(\lambda)\eta(\lambda)\,d\lambda. \end{aligned}$$

By the assumptions it suffices to show that

$$x_n \int_{-\pi}^{\pi} R_n(\lambda)\eta(\lambda)\,d\lambda = o_P(1).$$

Using the boundedness of η, the definition of $R_n(\lambda)$ (see Lemma 5.1) and Hölder's inequality, we find that, for some $c_1, c_2 > 0$,

$$\begin{aligned}
&\left|\int_{-\pi}^{\pi} R_n(\lambda)\eta(\lambda)\,d\lambda\right| \\
&\quad \le c_1 \int_{-\pi}^{\pi} |R_n(\lambda)|\,d\lambda \\
&\quad \le c_2 \left\{ \left(\int_{-\pi}^{\pi} I_{n,Z}(\lambda)\,d\lambda\right)^{1/2} \left(\int_{-\pi}^{\pi} |Y_n(\lambda)|^2\,d\lambda\right)^{1/2} + \int_{-\pi}^{\pi} |Y_n(\lambda)|^2\,d\lambda \right\}.
\end{aligned}$$

Thus it remains to show that

$$x_n^2 \int_{-\pi}^{\pi} |Y_n(\lambda)|^2\,d\lambda \to_{\mathscr{P}} 0.$$

We have

$$\begin{aligned}
\int_{-\pi}^{\pi} |Y_n(\lambda)|^2\,d\lambda \le cn^{-2/\alpha} \Bigg\{ &\int_{-\pi}^{\pi} \left| \sum_{j>n} \psi_j e^{-i\lambda j} \sum_{t=1-j}^{n-j} Z_t e^{-i\lambda t} \right|^2 d\lambda \\
&+ \int_{-\pi}^{\pi} \left| \sum_{j>n} \psi_j e^{-i\lambda j} \sum_{t=1}^{n} Z_t e^{-i\lambda t} \right|^2 d\lambda \\
&+ \int_{-\pi}^{\pi} \left| \sum_{j=1}^{n} \psi_j e^{-i\lambda j} \sum_{t=1-j}^{0} Z_t e^{-i\lambda t} \right|^2 d\lambda \\
&+ \int_{-\pi}^{\pi} \left| \sum_{j=1}^{n} \psi_j e^{-i\lambda j} \sum_{t=n-j+1}^{n} Z_t e^{-i\lambda t} \right|^2 d\lambda \Bigg\} \\
:= cn^{-2/\alpha}(V_1 + V_2 + V_3 + V_4).&
\end{aligned}$$

It suffices to show that the V_i are stochastically bounded. We will show this for V_1. The other estimates are similar. Note that $V_1 = \int_{-\pi}^{\pi} |Q(\lambda)|^2\,d\lambda$, where

$$Q(\lambda) = \sum_{j=-\infty}^{-1} Z_j \sum_{t=(n+1)\vee(1+j)}^{n-j} \psi_t e^{-i\lambda(t+j)}.$$

Let B_1 and B_2 be two independent Brownian motions on $[-\pi, \pi]$ and suppose that they are independent of $\{Z_t\}_{t \in \mathscr{Z}}$. Then

$$\begin{aligned}
&E \exp\left(ir \left(\int_{-\pi}^{\pi} \operatorname{Re}(Q(\lambda))\, dB_1(\lambda) + \int_{-\pi}^{\pi} \operatorname{Im}(Q(\lambda))\, dB_2(\lambda) \right) \right) \\
&= E\left(E\left(\exp\left(ir \left(\int_{-\pi}^{\pi} \operatorname{Re}(Q(\lambda))\, dB_1(\lambda) + \int_{-\pi}^{\pi} \operatorname{Im}(Q(\lambda))\, dB_2(\lambda) \right) \right) \middle| (Z_t) \right) \right) \\
&= E\left(E\left(\exp\left(ir \left(\left(\int_{-\pi}^{\pi} (\operatorname{Re}(Q(\lambda)))^2\, d\lambda \right)^{1/2} N_1 \right.\right.\right.\right. \\
&\qquad \left.\left.\left.\left. + \left(\int_{-\pi}^{\pi} (\operatorname{Im}(Q(\lambda)))^2\, d\lambda \right)^{1/2} N_2 \right) \right) \middle| (Zt) \right) \right) \\
&= E \exp\left(-\frac{r^2}{2} \int_{-\pi}^{\pi} |Q(\lambda)|^2\, d\lambda \right) \\
&= E \exp\left(-\frac{r^2}{2} V_1 \right).
\end{aligned}$$

Here N_1, N_2 are iid standard Gaussian random variables independent of $\{Z_t\}_{t \in \mathscr{Z}}$. It therefore suffices to prove that the real and the imaginary parts of $\int_{-\pi}^{\pi} Q(\lambda)\, dB_1(\lambda)$ are stochastically bounded. We restrict ourselves to the real part. We introduce the gauge function Λ_α for any random variable A by

$$\Lambda_\alpha(A) = \left(\sup_{t > 0} t^\alpha P(|A| > t) \right)^{1/\alpha}.$$

Then for any sequence $(d_i)_{i \in \mathscr{N}}$ of real numbers, we have for some constant $c_\alpha > 0$,

$$\Lambda_\alpha^\alpha \left(\sum_{i=1}^n d_i Z_i \right) \le c_\alpha \sum_{i=1}^n |d_i|^\alpha \Lambda_\alpha^\alpha (Z_0) \tag{6.4}$$

[see, e.g., Klüppelberg and Mikosch (1993), Lemma 3.4]. Write

$$D_j = \sum_{t=(n+1)\vee(1-j)}^{n-j} \psi_t \int_{-\pi}^{\pi} \operatorname{Re}(e^{-i\lambda(t+j)})\, dB_1(\lambda).$$

Then, for fixed $\varepsilon > 0$,

$$\begin{aligned}
&P\left(\left| \int_{-\pi}^{\pi} \operatorname{Re}(Q(\lambda))\, dB_1(\lambda) \right| \ge \varepsilon \right) \\
&\quad = P\left(\left| \sum_{j=-\infty}^{-1} Z_j D_j \right| \ge \varepsilon \right) \\
&\quad \le \varepsilon^{-\alpha} E\left(\sup_{s>0} s^\alpha P\left(\left| \sum_{j=-\infty}^{-1} Z_j D_j \right| > s \middle| B_1 \right) \right).
\end{aligned}$$

An application of inequality (6.4) yields that

$$
\begin{aligned}
P&\left(\left|\int_{-\pi}^{\pi} \operatorname{Re}(Q(\lambda))\, dB_1(\lambda)\right| \geq \varepsilon\right) \\
&\leq \varepsilon^{-\alpha} E\left(c_\alpha \sum_{j=-\infty}^{-1} \Lambda_\alpha^\alpha(Z_0) |D_j|^\alpha\right) \\
&= \varepsilon^{-\alpha} c_\alpha \Lambda_\alpha^\alpha(Z_0) \sum_{j=-\infty}^{-1} E\left|\sum_{t=(n+1)\vee(1-j)}^{n-j} \psi_t \int_{-\pi}^{\pi} \operatorname{Re}(e^{-i\lambda(t+j)})\, dB_1(\lambda)\right|^\alpha \\
&\leq \varepsilon^{-\alpha} c_1 \Lambda_\alpha^\alpha(Z_0) \sum_{j=-\infty}^{-1} E|N_j|^\alpha \left(\sum_{t=(n+1)\vee(1-j)}^{n-j} |\psi_t|^2\right)^{\alpha/2} \\
&\leq \varepsilon^{-\alpha} c_2 \sum_{j=1}^{\infty} |\psi_j|^{\alpha\wedge 1} j.
\end{aligned}
$$

Here (N_j) is a sequence of identically distributed Gaussian random variables and c_1, c_2 are positive constants. Choosing ε large enough then establishes the stochastic boundedness of V_1. □

LEMMA 6.3. *Suppose the assumptions of Theorem 2.2 hold. Furthermore, let η be an even continuous function on $[-\pi, \pi]$ and let the Fourier coefficients f_k, $k \in \mathscr{Z}$, of $\eta(\lambda)g(\lambda, \beta_0)$ satisfy $f_0 = 0$ and $\sum_{k=-\infty}^{\infty} |f_k|^\mu < \infty$ for some $\mu \in (0, 1 \wedge \alpha)$. Then*

$$
\left(\frac{n}{\ln n}\right)^{1/\alpha} \int_{-\pi}^{\pi} \tilde{I}_{n,X}(\lambda)\eta(\lambda)\, d\lambda \to_{\mathscr{D}} 4\pi \Psi^{-2}(\beta_0) \sum_{k=1}^{\infty} \frac{Y_k}{Y_0} f_k,
$$

where $Y_0, Y_1, Y_2, \ldots$ are independent random variables, Y_0 is positive $\alpha/2$-stable and $(Y_t)_{t\in\mathscr{N}}$ are iid symmetric α-stable with characteristic function $E\exp(itY_1) = \exp(-C_\alpha |t|^\alpha)$, $t \in \mathscr{R}$.

PROOF. In view of Lemma 6.2 it suffices to show that

$$
x_n \int_{-\pi}^{\pi} \tilde{I}_{n,Z}(\lambda)\eta(\lambda) g(\lambda, \beta_0)\, d\lambda \to_{\mathscr{D}} 4\pi \sum_{k=1}^{\infty} \frac{Y_k}{Y_0} f_k,
$$

where $x_n = (n/\ln n)^{1/\alpha}$. Set

$$
\chi(\lambda) = \eta(\lambda) g(\lambda, \beta_0)
$$

and

$$
\chi_m(\lambda) = \sum_{|k|\leq m} f_k e^{i\lambda k} \quad \text{with } f_k = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-i\lambda k}\chi(\lambda)\, d\lambda.
$$

The assumptions on (f_k) imply that

$$\chi_m(\lambda) \to \chi(\lambda) = \sum_{k=-\infty}^{\infty} f_k e^{i\lambda k}, \qquad m \to \infty,$$

uniformly in λ. We show that for all $\varepsilon > 0$,

$$\lim_{m\to\infty} \limsup_{n\to\infty} P\left(x_n \left| \int_{-\pi}^{\pi} \tilde{I}_{n,Z}(\lambda)(\chi(\lambda) - \chi_m(\lambda))\, d\lambda \right| > \varepsilon \right) = 0. \tag{6.5}$$

For $n \in \mathscr{N}$ we set $y_n = (n \ln n)^{1/\alpha}$. Then for $n > m$ there exists some $c_1 > 0$ such that

$$\begin{aligned}
V_1 &= x_n \int_{-\pi}^{\pi} \tilde{I}_{n,Z}(\lambda)(\chi(\lambda) - \chi_m(\lambda))\, d\lambda \\
&= x_n \int_{-\pi}^{\pi} \left(\sum_{|h|<n} \tilde{\gamma}_{n,Z}(h) e^{-i\lambda h} \sum_{|k|>m} f_k e^{i\lambda k} \right) d\lambda \\
&= x_n 2\pi \sum_{m<|h|<n} \tilde{\gamma}_{n,Z}(h) f_h \\
&= c_1 \gamma_{n,Z}^{-2} y_n^{-1} \left\{ \sum_{h=m+1}^{n-1} f_h \sum_{t=1}^{n-h} Z_t Z_{t+h} \right\} \\
&= c_1 \gamma_{n,Z}^{-2} y_n^{-1} \sum_{t=1}^{n-m-1} Z_t \sum_{h=m+1}^{n-t} f_h Z_{t+h} \\
&= c_1 \gamma_{n,Z}^{-2} y_n^{-1} \sum_{t=1}^{n-m-1} Z_t \sum_{h=m+t+1}^{n} f_{h-t} Z_h \\
&:= c_1 \gamma_{n,Z}^{-2} V_2.
\end{aligned}$$

Since $\gamma_{n,Z}^2 \to_{\mathscr{D}} Y_0$ for some positive $\alpha/2$-stable random variable Y_0, distributed as in the statement of the lemma, (6.5) will follow once we show that for every $\varepsilon > 0$,

$$\lim_{m\to\infty} \limsup_{n\to\infty} P(|V_2| > \varepsilon) = 0.$$

An application of Theorem 3.1 of Rosinski and Woyczynski (1987) yields for some $c_2 > 0$ depending on ε that

$$P(|V_2| > \varepsilon) \le c_2 \frac{1}{n} \sum_{t=1}^{n-m-1} \sum_{h=m+t+1}^{n} |f_{h-t}|^{\alpha} \left(1 + \log^+ \frac{1}{|f_{h-t}|} \right),$$

where $\log^+ x = \max(0, \ln x)$. Note that for $x \in (0,1)$,

$$x^{\alpha} \left(1 + \log^+ \frac{1}{x} \right) \le x^{\mu},$$

where $\mu \in (0, 1 \wedge \alpha)$. Hence for constants $c_2, c_3 > 0$, both depending on ε,

$$P(|V_2| > \varepsilon) \le c_2 \frac{1}{n} \sum_{t=1}^{n-m-1} \sum_{h=m+t+1}^{n} |f_{h-t}|^{\mu}$$
$$\le c_3 \frac{1}{n} \sum_{l=m+1}^{n} (n-l)|f_l|^{\mu} \le c_3 \sum_{l=m+1}^{\infty} |f_l|^{\mu}$$

and, by the assumptions, the rhs converges to 0 as $m \to \infty$. This proves (6.5). Now it remains to show that, for each fixed m,

$$V_3 = x_n \int_{-x}^{\pi} \tilde{I}_{n,Z}(\lambda) \chi_m(\lambda)\, d\lambda \to_{\mathscr{D}} 2\pi \sum_{|k| \le m} f_k \frac{Y_k}{Y_0}, \tag{6.6}$$

as $n \to \infty$. For $n > m$ we have

$$V_3 = x_n \int_{-\pi}^{\pi} \left(\sum_{|h|<n} \tilde{\gamma}_{n,Z}(h) e^{-i\lambda h} \sum_{|k| \le m} f_k e^{i\lambda k} \right) d\lambda$$
$$= x_n 2\pi \sum_{|h| \le m} \tilde{\gamma}_{n,Z}(h) f_h$$
$$= 2\pi \gamma_{n,Z}^{-2} \sum_{|h| \le m} f_h \left(y_n^{-1} \sum_{t=1}^{n-|h|} Z_t Z_{t+|h|} \right).$$

Theorem 3.3 of Davis and Resnick (1986) gives, for $h > 0$,

$$\left(\gamma_{n,Z}^2, y_n^{-1} \sum_{t=1}^{n-1} Z_t Z_{t+1}, \dots, y_n^{-1} \sum_{t=1}^{n-h} Z_t Z_{t+h} \right) \to_{\mathscr{D}} (Y_0, Y_1, \dots, Y_h).$$

The specific scaling constants in the statement of the lemma, and in Theorem 5.2, then follow from the representation of the Y_i given in Davis and Resnick (1986) and the results of Le Page (1980).

This together with the continuous mapping theorem and the fact that $f_0 = 0$ proves (6.6) and the lemma. □

PROOF OF THEOREM 2.2. A Taylor expansion about β_0 gives

$$\begin{aligned} \frac{\partial \sigma_n^2(\beta_0)}{\partial \beta} &= \frac{\partial \sigma_n^2(\beta_n)}{\partial \beta} - (\beta_n - \beta_0) \frac{\partial^2 \sigma_n^2(\beta_n^*)}{\partial \beta^2} \\ &= -(\beta_n - \beta_0) \frac{\partial^2 \sigma_n^2(\beta_n^*)}{\partial \beta^2} \end{aligned} \tag{6.7}$$

for some β_n^* with $\| \beta_n^* - \beta_n \| \le \| \beta_n - \beta_0 \|$, where $\| \cdot \|$ denotes the Euclidean norm. Now

$$\frac{\partial^2 \sigma_n^2(\beta_n^*)}{\partial \beta^2} = \int_{-\pi}^{\pi} \tilde{I}_{n,X}(\lambda) \frac{\partial^2 g^{-1}(\lambda, \beta_n^*)}{\partial \beta^2}\, d\lambda$$

and since $\beta_n^* \to_{\mathscr{P}} \beta_0$, similar arguments as in the proof of Lemma 6.1 yield that

$$\frac{\partial^2 \sigma_n^2(\beta_n^*)}{\partial \beta^2} \to_{\mathscr{P}} \Psi^{-2}(\beta_0) \int_{-\pi}^{\pi} g(\lambda, \beta_0) \frac{\partial^2 g^{-1}(\lambda, \beta_0)}{\partial \beta^2}\, d\lambda.$$

Following the lines of the proof in Brockwell and Davis [(1991) after (10.8.39)], the same arguments lead to

$$\frac{\partial^2 \sigma_n^2(\beta_n^*)}{\partial \beta^2} \to_{\mathscr{P}} \Psi^{-2}(\beta_0) W(\beta_0),$$

where W is the matrix defined in the statement of the theorem. Hence it suffices to show that

$$x_n \frac{\partial \sigma_n^2(\beta_0)}{\partial \beta} \to_{\mathscr{D}} 4\pi \Psi^{-2}(\beta_0) \sum_{k=1}^{\infty} \frac{Y_k}{Y_0} b_k, \tag{6.8}$$

where b_k is defined by (2.5). [Note that we may ignore the negative sign on the right-hand side of (6.7) because of the symmetry of the limit distribution.] Equivalently, by the Cramér–Wold device, it suffices to show that for all vectors $c \in \mathscr{R}^{p+q}$,

$$x_n c^T \frac{\partial \sigma_n^2(\beta_0)}{\partial \beta} \to_{\mathscr{D}} 4\pi \Psi^{-2}(\beta_0) \sum_{k=1}^{\infty} \frac{Y_k}{Y_0} c^T b_k.$$

We have

$$\begin{aligned} c^T \frac{\partial \sigma_n^2(\beta_0)}{\partial \beta} &= c^T \int_{-\pi}^{\pi} \tilde{I}_{n,X}(\lambda) \frac{\partial g^{-1}(\lambda, \beta_0)}{\partial \beta}\, d\lambda \\ &:= \int_{-\pi}^{\pi} \tilde{I}_{n,X}(\lambda) \eta(\lambda)\, d\lambda, \end{aligned}$$

where

$$\eta(\lambda) = c^T \frac{\partial g^{-1}(\lambda, \beta_0)}{\partial \beta}$$

is an even continuous function. Furthermore, it is not difficult to see that the Fourier coefficients of $\eta(\lambda) g(\lambda, \beta_0)$ satisfy the conditions of Lemma 6.3. An application of this lemma implies that

$$x_n \int_{-\pi}^{\pi} \tilde{I}_{n,X}(\lambda) \eta(\lambda)\, d\lambda \to_{\mathscr{D}} 4\pi \Psi^{-2}(\beta_0) \sum_{k=1}^{\infty} \frac{Y_k}{Y_0} f_k,$$

where f_k are the Fourier coefficients of $\eta(\lambda) g(\lambda, \beta_0)$; that is,

$$f_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-ik\lambda} c^T \frac{\partial g^{-1}(\lambda, \beta_0)}{\partial \beta} g(\lambda, \beta_0)\, d\lambda.$$

Since $f_k = c^T b_k$, this implies (6.8) and so the theorem. □

Acknowledgments. CK and TM take pleasure in thanking Professor H.-R. Künsch for his everyday readiness to discuss the topic of this paper. All four authors thank both the Associate Editor and a referee for a number of useful comments.

REFERENCES

AN, H. Z. and CHEN, Z. G. (1982). On convergence of LAD estimates in autoregression with infinite variance. *J. Multivariate Anal.* **12** 335–345.

BHANSALI, R. J. (1984). Order determination for processes with infinite variance. In *Robust and Nonlinear Time Series Analysis* (J. Franke, W. Härdle and D. Martin, eds.) 17–25. Springer, New York.

BHANSALI, R. J. (1988). Consistent order determination for processes with infinite variance. *J. Roy. Statist. Soc. Ser. B* **50** 46–60.

BINGHAM, N. H., GOLDIE, C. M. and TEUGELS, J. L. (1987). *Regular Variation*. Cambridge Univ. Press.

BROCKWELL, P. J. and DAVIS, R. A. (1991). *Time Series: Theory and Methods*, 2nd ed. Springer, Berlin.

CHAMBERS, J. M., MALLOWS, C. L. and STUCK, B. W. (1976). A method for simulating stable random variables. *J. Amer. Statist. Assoc.* **71** 340–344.

DAHLHAUS, R. (1989). Efficient parameter estimation for self-similar processes. *Ann. Statist.* **17** 1749–1766.

DAVIS, R. A., KNIGHT, K. and LIU, J. (1992). *M*-estimation for autoregressions with infinite variance. *Stochastic Process. Appl.* **40** 145–180.

DAVIS, R. A. and RESNICK, S. I. (1986). Limit theory for the sample covariance and correlation functions of moving averages. *Ann. Statist.* **14** 533–558.

FEIGIN, P. and RESNICK, S. I. (1992). Estimation for autoregressive processes with positive innovations. *Comm. Statist. Stochastic Models* **8** 479–498.

FEIGIN, P. and RESNICK, S. (1994). Limit distributions for linear programming time series estimators. *Stochastic Process. Appl.* **51** 135–166.

FELLER, W. (1971). *An Introduction to Probability Theory and Its Applications*, **2** 2nd ed. Wiley, New York.

FOX, R. and TAQQU, M. S. (1986). Large sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *Ann. Statist.* **14** 517–532.

GADRICH, T. (1993). Parameter estimation for ARMA processes with symmetric stable innovations. D.Sc. Thesis, Technion, Haifa (in Hebrew).

HAHN, M. G. and WEINER, D. C. (1991). On joint estimation of an exponent of regular variation and an asymmetry parameter for tail distributions. In *Sums, Trimmed Sums and Extremes* (M. G. Hahn, D. M. Mason and D. C. Weiner, eds.). *Progress in Probability* **23**. Birkhäuser, Boston.

HANNAN, E. J. (1973). The asymptotic theory of linear time-series models. *J. Appl. Probab.* **10** 130–145.

HANNAN, E. J. and KANTER, M. (1977). Autoregressive processes with infinite variance. *J. Appl. Probab.* **14** 411–415.

HSING, T. (1991). On tail index estimation using dependent data. *Ann. Statist.* **19** 1547–1569.

HSING, T. (1993). On some estimates based on sample behavior near high level excursions. *Probab. Theory Related Fields* **95** 331–356.

KLÜPPELBERG, C. and MIKOSCH, T. (1993). Spectral estimates and stable processes. *Stochastic Process. Appl.* **47** 323–344.

KLÜPPELBERG, C. and MIKOSCH, T. (1994). Some limit theory for the self-normalised periodogram of p-stable processes. *Scand. J. Statist.* **21** 485–491.

KNIGHT, K. (1989). Consistency of Akaike's information criterion for infinite autoregressive processes. *Ann. Statist.* **17** 824–840.

KOUTROUVELIS, I. A. (1980). Regression-type estimation of the parameters of stable laws. *J. Amer. Statist. Assoc.* **75** 918–928.

LE PAGE, R. (1980). Multidimensional infinitely divisible variables and processes. Part I: Stable case. *Probability Theory on Vector Spaces IV. Lecture Notes in Math.* **1391** 153–163. Springer, Berlin.

MCCULLOCH, J. H. (1986). Simple consistent estimators of stable distribution parameters. *Comm. Statist. Simulation Comput.* **15** 1109–1136.

ROSINSKI, J. and WOYCZYNSKI, W. A. (1987). Multilinear forms in Pareto-like random variables and product random measures. *Colloq. Math.* **51** 303–313.

WHITTLE, P. (1953). Estimation and information in stationary time series. *Ark. Mat.* **2** 423–434.

THOMAS MIKOSCH
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF GRONINGEN
NL-9700 AV GRONINGEN
THE NETHERLANDS

TAMAR GADRICH
ROBERT J. ADLER
FACULTY OF INDUSTRIAL ENGINEERING
AND MANAGEMENT
TECHNION—ISRAEL INSTITUTE OF TECHNOLOGY
HAIFA, 32000
ISRAEL

CLAUDIA KLÜPPELBERG
DEPARTMENT OF MATHEMATICS
ETH ZÜRICH
CH-8092 ZÜRICH
SWITZERLAND